Knowledge and Information Systems (2025) 67:4349–4371
https://doi.org/10.1007/s10115-025-02363-z

REGULAR PAPER


# Spatial homogeneity-aware transfer learning for urban flow prediction

**Yinghui Liu[1] · Guojiang Shen[1] · Yanjie Fu[2] · Zehui Feng[1] · Zhenzhen Zhao[1] · Xiangjie Kong[1]**

Received: 28 November 2023 / Revised: 20 August 2024 / Accepted: 2 February 2025 /
Published online: 24 February 2025


**Abstract**
The precondition for a deep prediction model to achieve high prediction accuracy is sufficient data, which is not always available in practice. Thus, urban transfer learning methods (i.e., fine-tuning-based methods) are successively proposed to mitigate this issue. However, these existing approaches do not estimate source knowledge transferability and therefore easily lead to negative transfer. Spatial homogeneity (i.e., F1 score gained by link prediction) can provide fine-grained topological structure indication for source knowledge transfer. To this end, we propose a spatial homogeneity-aware transfer learning framework named SHTL for urban flow prediction. In particular, SHTL consists of a link prediction model and an urban flow prediction model. Firstly, the link prediction model is used to capture regional road network topology and regional spatial homogeneity is obtained by evaluating model predictability in each region. Secondly, the urban flow prediction model is optimized by selective source training and target fine-tuning based on spatial homogeneity. We evaluate SHTL on real-world taxi and bike datasets and the result shows that SHTL outperforms the state-of-the-art baselines.

**Keywords** Urban flow prediction · Spatial-temporal data · Spatial homogeneity · Transfer learning

✉ Guojiang Shen
gjshen1975@zjut.edu.cn

✉ Xiangjie Kong
xjkong@ieee.org

Yinghui Liu
LiuYingHui240@outlook.com

Yanjie Fu
yanjie.fu@asu.edu

Zehui Feng
202003151208@zjut.edu.cn

Zhenzhen Zhao
zhenzhenzhao_97@outlook.com

[1] College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, Zhejiang, China

[2] School of Computing and AI, Arizona State University, Tempe, AZ 85287, USA

## 1 Introduction

With the rapid iteration of city infrastructures (e.g., traffic monitoring devices, air quality monitoring stations) and the popularity of various mobile devices and transportation, huge amounts of urban data such as taxicab GPS data and meteorology data keep emerging. Spatiotemporal data-based prediction (including traffic flow prediction [20, 23, 31, 44], traffic event prediction [12, 15, 18, 27], crowd flow prediction [22, 24, 30], and so on), as a key type of task in urban computing, has therefore been researched extensively. In the past years, diverse deep learning models such as recurrent neural networks (RNNs), convolutional neural networks (CNNs), and graph neural networks (GNNs) [14] have shown great performance on these tasks. However, there is a critical problem in that these models require large-scale urban data but data in some cities tend to be scarce due to the data collection problem [39, 41], data privacy issues [25], or other reasons.

To alleviate the above issue, an effective method is actually to use transfer learning [33], which can transfer helpful knowledge from one domain to another one. Existing works such as [40, 42, 43] have successfully achieved cross-city crow flow prediction, mobile traffic prediction, and other tasks via fine-tuning-based model-transfer (parameter-transfer) approach. The fine-tuning-based model-transfer approach includes two steps: 1) train the model with abundant source data to capture knowledge from the source domain (source training); 2) train the model transferred from the source domain with scarce target data so that the target domain benefits from source domain knowledge (target fine-tuning).

However, previous works ignore source knowledge transferability during transfer, which is prone to negative transfer [16]. Besides, they only focus on the spatiotemporal correlation of city-specific region features (for example, crowd flow) during the target fine-tuning without taking the potential topological similarities of road network units into account. Recently, a study [46] attracts our attention, which quantifies the spatial homogeneity of urban road networks (URNs) via GNNs. It is proved that the provided indicator has strong correlations with most road network measurement metrics (e.g., number of nodes [3], betweenness [34], logarithmic circuity [10], and the fraction of bridge edges [2]) and reveals both intracity socioeconomic development (for example, the growth of GDP and population) and intercity historical transfer (for example, some American urban plannings are deeply impacted by Europe). The spatial homogeneity, as the F1 score gained by link prediction [21], not only means the topological similarity between each region (subnetwork) and the entire urban road network in the same city (i.e., intracity spatial homogeneity) but also means the topological similarity between each region (subnetwork) in the target city and the entire road network in source city (i.e., intercity spatial homogeneity). The visualization of the spatial homogeneity can be found in Sect. 4.2. Since the strong dependence of socioeconomic activities on URNs, measuring the spatial homogeneity of URNs can assist in the evaluation of source knowledge transferability.

In this article, we first propose a novel **S**patial **H**omogeneity-aware **T**ransfer **L**earning framework called SHTL for urban flow prediction in which we estimate the knowledge transferability of source region according to spatial homogeneity value. SHTL comprises two tasks, link prediction and urban flow prediction based on fine-tuning. Specifically, we use a link prediction model $F_{\theta_L}$ to capture the topological structure of each regional road network in the target domain. By evaluating the performance of link prediction model $F_{\theta_L}$ testing in the source domain, we gain the F1 score of each source region (i.e., spatial homogeneity value). Then, we apply source regional spatial homogeneity values in urban flow prediction based on fine-tuning, where we take source regional spatial homogeneity values $M_S$ as

weights and assign them to source regions during both processes. More homogeneous source regions are assigned higher weights. During the source training, an urban flow prediction model $F_\theta$ is selectively trained on the source domain via $M_S$ to initialize target fine-tuning. Before the target fine-tuning, we match similar region pairs via $M_S$ and data similarities. By minimizing the discrepancy between region representations of similar regions during the target fine-tuning, we finally obtain the optimized urban flow prediction model $F_\theta$. Extensive experiments on real-world taxi and bike datasets are performed and the results show that SHTL outperforms the state-of-the-art baselines under the same settings.

Briefly, our major contributions can be summarized as follows:

- We propose the framework SHTL to achieve urban flow prediction in a data-scarce city, in which we estimate source knowledge transferability via the spatial homogeneity of URNs. To our knowledge, this is the first work that explores the influence of the topologies of URNs on transfer learning for urban flow prediction.
- We develop selective source training based on spatial homogeneity to provide transferability for source knowledge. In addition, we design a matching function based on spatial homogeneity to make the urban flow prediction model accurately learn similar cross-domain regional knowledge during target fine-tuning.
- We perform extensive experiments on real-world taxi and bike datasets to validate and explain the effectiveness of the spatial homogeneity of URNs.

The remainder of this article is organized as follows. Section 2 will review related works. Section 3 will provide corresponding definitions and problem formulations. We will show the overall framework and detail its components in Sect. 4. Experimental results will be presented in Sect. 5, followed by the conclusion in Sect. 6.

## 2 Related work

In this section, owing to the high relevance of our work on grid-based urban flow prediction, transfer learning for urban flow prediction, and urban road network topology feature, we will review the related works from these three aspects.

### 2.1 Grid-based urban flow prediction

With recent advances in ubiquitous devices, spatiotemporal data prediction has been an appealing domain in urban computing. The Geographical Information System (GIS) usually divides a city into multiple grids (a grid represents a region) for region-level traffic analysis and prediction such as traffic congestion [11] and population flow [9]. Compared with the traditional statistics-based methods (e.g., support vector regression (SVR) and ARIMA), deep learning methods have excellent learning capacity to capture the complex spatiotemporal dependencies of urban data. Thus, the first CNN-based architecture called DeepST was presented by [49] to predict the grid-based urban flows. Then, they further presented STResNet [50] to forecast both inflow and outflow of crowds in every grid unit. However, these models only simply treat the information in the same time intervals as multiple channels of convolution according to temporal properties (i.e., closeness, period, and trend) without taking into account the temporal dependence in the intermediate layers. To this end, many studies [6, 7, 38] capture the dynamic spatiotemporal relationships by combining the CNNs with RNNs. However, the above methods undergo accuracy degradation when data is sparse, which is the problem we solve in this paper.

### 2.2 Transfer learning for urban flow prediction

In the case of data scarcity, how to obtain a remarkable city-wide prediction model is a question that is worthy of exploration. Transfer learning, as a successful approach in both Computer Vision (CV) [48] and Natural Language Processing (NLP) [4], can be also applied to urban flow prediction. Urban transfer learning is a very promising tool that can transfer some prior knowledge from the source city to the target city. Consequently, some fine-tuning-based cross-city transfer learning models including RegionTrans [40], MetaST [47], and ST-DAAN [42] have been presented to predict urban flow in the data-scarce city. RegionTrans transfers a well-trained model on the source city to the target city and then fine-tuning by data similarity (i.e., Pearson coefficient); MetaST captures the common long-term temporal features of the spatiotemporal data from multiple cities and then transfers them to the target city; ST-DAAN fine-tunes the well-trained model on the source city by leveraging deep adaptation networks (DAN) to fine-tuning. Recently, STAN [8] employs both temporal domain adaptation and spatial domain adaptation to enhance spatiotemporal knowledge transfer. ST-GFSL [29] introduces spatiotemporal graph few-shot learning to transfer cross-city meta-knowledge for non-grid flow prediction. CrossTReS [16] selectively learns from the source city to adapt to target fine-tuning by learning generalizable regional functional features (e.g., Points of Interest (POIs)) between two cities. Different from existing works, we utilize effectively spatial homogeneity of URNs on both intracity and intercity transfer learning for urban flow prediction.

### 2.3 Urban road network topology feature

Traditional road network topology features can be classified into two categories. The first type is the global network features such as "scaling laws" [37, 51] which expound on universal mechanisms of road networks. The second type is the local network features that capture multi-hop relationships by measuring the "local function" of a node in the network including the local version of the original global features (e.g., local betweenness [17, 32] and local closeness [5]) and subgraph theories features (e.g., motifs [1]). Both types measure the property of a road network by a specific value but do not weigh up the similarity between road networks. To measure the similarity index, we still need an extra process consisting of vector normalization rules (for example, normalize by the minimal and maximal values) and similarity distance definitions (for example, the Euclidean distance). Diverse combinations make the final similarity index to be very uncertain and lack interpretability. Hence, the spatial homogeneity metric quantified by GNNs which captures the nonlinear structural correlations of road networks at both micro- and macro-level has been presented to directly measure the similarity of two subnetworks. In this paper, we aim to offer knowledge transferability to urban transfer learning by using the explainable spatial homogeneity metric.

## 3 Preliminary

In this section, we will first provide the necessary definitions to help us state the studied problem. Then, the correlative problem formulations including link prediction and urban flow prediction will be given.

### 3.1 Link prediction

**Definition 1** [Urban Regions] Based on the longitude and latitude, a city $\mathcal{C}$ can be partitioned into a grid map with totally $N = W_{\mathcal{C}} \times H_{\mathcal{C}}$ equal-size grids (e.g., 1 km×1 km). Each grid stands for a city region $r_{[w,h]}$ where $[w, h]$ is the region coordinate in the grid map and the set of all regions in city $\mathcal{C}$ is also briefly defined as a matrix $\mathcal{C} \in \mathbb{R}^{W_{\mathcal{C}} \times H_{\mathcal{C}}}$.

**Definition 2** [Regional Road Network Graphs] The road network in region r is modeled as the graph $\mathcal{G}_r = (\mathcal{V}_r, \mathcal{E}_r) \subset \mathcal{G}_{\mathcal{C}}$, where $\mathcal{V}_r$ and $\mathcal{E}_r$ are the sets of nodes and edges, respectively. Intersections and road segments in this region are defined as nodes $v_i \in \mathcal{V}_r$ and edges (links) $(v_i, v_j) \in \mathcal{E}_r$.

According to the above definitions, we present the statement of link prediction. In this paper, we focus on predicting the link of every grid (region).

**Problem 1** (Link Prediction) Given a connected but incomplete road network graph $\mathcal{G}_r{}' = \left(\mathcal{V}_r, \mathcal{E}_r{}'\right)$, where $\mathcal{E}_r{}' \subseteq \mathcal{E}_r$, our target is to predict the full set of $\mathcal{E}_r$ by computing the connecting strength of every possible node pair that is not directly connected by a link in $\mathcal{E}_r{}'$.

### 3.2 Transfer learning for urban flow prediction

**Definition 3** (Urban Flow) We divide the time range in city $\mathcal{C}$ into equal-length time intervals $t = 1, 2, ...T_{\mathcal{C}}$. Then, the urban flow can be denoted as a 4-dimension tensor $X_{\mathcal{C}} \in \mathbb{R}^{T_{\mathcal{C}} \times K \times W_{\mathcal{C}} \times H_{\mathcal{C}}}$, where $K$ is the number of flow measurements (i.e., inflow/outflow) and each entry $x_{r,k}^{(t)}$ denotes the value of the $k$-th measurement at a certain time $t$ in region $r$.

We then formulate the problem of transfer learning for urban flow prediction as follows. Note that the method of transfer learning is based on fine-tuning, the same as the existing works [8, 16, 29, 40, 42, 47].

**Problem 2** (Transfer Learning for Urban Flow Prediction) Given a source domain $\mathcal{S}$ and a target domain $\mathcal{T}$ with flow data $X_{\mathcal{S}} = \left\{x_r^{(t)} \mid t \in [1, T_{\mathcal{S}}], r \in \mathcal{S}\right\}$, $X_{\mathcal{T}} = \left\{x_r^{(t)} \mid t \in [1, T_{\mathcal{T}}], r \in \mathcal{T}\right\}$, $T_{\mathcal{S}} \gg T_{\mathcal{T}}$, we first train the prediction model $F_{\theta_{\mathcal{S}}}$ with abundant source domain data $X_{\mathcal{S}}$ in **source training**:

$$\min_{\theta_{\mathcal{S}}} \mathcal{L}_{\mathcal{S}}\left(X_{\mathcal{S}}\right) = \sum_{r \in \mathcal{S}} \sum_{t=1}^{T_{\mathcal{S}}} L\left(\hat{x}_r^{(t)}, x_r^{(t)}\right), \text{ where } \hat{x}_r^{(t)} = F_{\theta_{\mathcal{S}}}\left(\left[x_r^{(t-\tau)}, ...x_r^{(t-1)}\right], X_{\mathcal{S}}\right). \quad (1)$$

$\hat{x}_r^{(t)}, x_r^{(t)}$ are the predicted and true values of source region $r$ at time step $t$. $L$ is the prediction error which can be an absolute error, root squared error, etc. We then transfer knowledge learned from the source domain to the target domain so that learn the prediction model $F_{\theta_{\mathcal{T}}}$ with target domain data in **target fine-tuning**:

$$\min_{\theta_{\mathcal{T}}} \mathcal{L}_{\mathcal{T}}\left(X_{\mathcal{T}}; \theta_{\mathcal{S}}\right) = \sum_{r \in \mathcal{T}} \sum_{t=1}^{T_{\mathcal{T}}} L\left(\hat{x}_r^{(t)}, x_r^{(t)}\right), \text{ where } \hat{x}_r^{(t)} = F_{\theta_{\mathcal{T}}}\left(\left[x_r^{(t-\tau)}, ...x_r^{(t-1)}\right], X_{\mathcal{T}}\right). \quad (2)$$

$\hat{x}_r^{(t)}, x_r^{(t)}$ are the predicted and true values of target region $r$ at time step $t$. $\mathcal{L}_{\mathcal{T}}\left(X_{\mathcal{T}}; \theta_{\mathcal{S}}\right)$ denotes model parameter $\theta_{\mathcal{T}}$ is initialized with $\theta_{\mathcal{S}}$. Finally, the problem is to gain the prediction model $F_\theta$ by the parameter $\theta_{\mathcal{T}}$ transfer. Note that the data of two domains can be collected from the same city or two different cities.

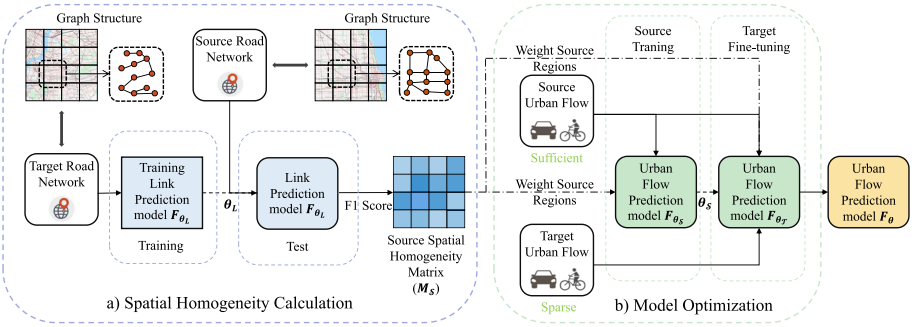



**Fig. 1** Overview of SHTL. The dotted line represents the parameter transfer and the chain line represents the action of weighting

## 4 Methodology

In this section, our proposed general framework SHTL is introduced. We first provide an overview of the framework and then introduce its components in detail.

### 4.1 Overview of SHTL

Figure 1 illustrates the framework SHTL, which consists of two major tasks: *Link Prediction for Spatial Homogeneity Calculation* and *Transfer Learning for Model (Parameter) Optimization*:

- **Link Prediction**. Link prediction is a process that predicts the existence of hidden links based on existing links. In urban computing, it can help us seek the hidden road segments between two intersections in a city and therefore indirectly reflect the topological structure of urban road networks. In the first task, to gain the spatial homogeneity matrix $M_{\mathcal{S}}$, we use road network data in the target domain to train $F_{\theta_L}$, test it with road network data in the source domain, and finally obtain the spatial homogeneity matrix values by taking the prediction accuracy metric F1 score (which is a standard and extensively used in machine-learning classification tasks with the range between 0 and 1). $M_{\mathcal{S}}$ means the topological similarities between all regions in the source domain and the global road network of the target domain. Note that when the source and target domain refer to the same city, we define the train and test set of the city as the source domain data and target domain data.
- **Transfer Learning**. In the second task, the spatial homogeneity matrix $M_{\mathcal{S}}$ measured from the first task is used to weight source regions to assist in the optimization of model $F_\theta$. As shown in Fig. 1, there are weightings in both source training and target fine-tuning. The first weighting is to select the helpful source regions [16] and the second weighting is for region matching [40] that matches these regions with both similar spatiotemporal data and road networks. All in all, $F_\theta$ is optimized in both source training with $M_{\mathcal{S}}$ and target fine-tuning with $M_{\mathcal{S}}$.

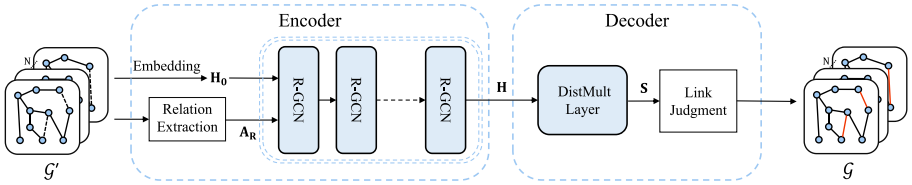



**Fig. 2** The detail of link prediction model $F_{\theta_L}$, where the black dotted line represents the masked link and the red line represents the predicted link. Specifically, $F_{\theta_L}$ consists of the encoder and the decoder. The node representations $\mathbf{H}$ generated by the encoder are regarded as the inputs for the decoder. The decoder utilizes the representations of two nodes to calculate the connecting strength $\mathbf{S}$ and predict the existence of the links

## 4.2 Spatial homogeneity calculation

In this subsection, the detailed pipeline of the link prediction model $F_{\theta_L}$ and the calculation process of the spatial homogeneity will be given.

### 4.2.1 Link prediction model $F_{\theta_L}$

GNN is a powerful deep learning model that can capture indirect neighborhood node relationships in contrast to existing URN metrics and the link prediction model $F_{\theta_L}$ is a GNN-based model. Figure 2 depicts the whole pipeline of $F_{\theta_L}$. First, urban road network data is modeled as graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $N$ subgraphs based on Def. 2. We hide some current links to produce the incomplete graph $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$. After embedding and relation extraction, we get the inputs of the GNN model: the original node representations (embeddings) $\mathbf{H_0}$ whose weights follow a standard normal distribution and relational adjacency matrix $\mathbf{A_R}$. Note that in the relation extraction, the link relation is defined from the relative angle: we evenly divide 360° into eight directions (i.e., north, south, west, east, northwest, northeast, southwest, and northwest); therefore, $\mathbf{A_R}$ includes eight types of relations. Following [46], we then choose relational graph convolutional network (RGCN) [35] that achieves the best performance in the link prediction of URNs to construct the node representations $\mathbf{H}$. The information propagation from layer $l$ to $(l+1)$ is:

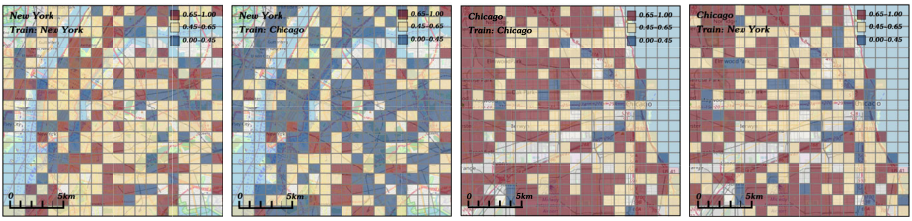
$$\mathbf{H}_i^{l+1} = tanh\left(\sum_{\mathbf{r}\in\mathbf{R}}\sum_{j\in\mathbf{A}_{\mathbf{r},i}}\frac{1}{c_i}\mathbf{W}_{\mathbf{r}}^l\mathbf{H}_j^l + \mathbf{W}_0^l\mathbf{H}_i^l\right), \tag{3}$$

where $\mathbf{H}_i^l, \mathbf{H}_i^{l+1}$ are the representations of node $i$ at layers $l$ and $l+1$, $\mathbf{r} \in \mathbf{R}$ is the link relation, $\mathbf{A}_{\mathbf{r},i}$ represents the adjacent node set of node $i$ under the relation $\mathbf{r}$. $c_i$ is a normalization constant and $\mathbf{W_r}$ is the learnable matrix. After the representations learning, $\mathbf{H}$ enters the DistMult layer and the connection strength is defined as:

$$\mathbf{S}(i, j) = \mathbf{H}_i^T\mathbf{W_d}\mathbf{H}_j, \tag{4}$$

where $\mathbf{H}_i, \mathbf{H}_j$ are the final representations of nodes $i$, $j$, $\mathbf{W_d}$ is the learnable diagonal matrix. We use the sigmoid function to transform the range of connection strength to (0,1):

$$\sigma\left(\mathbf{S}(i, j)\right) = \frac{1}{1 + e^{-\mathbf{S}(i,j)}}. \tag{5}$$




(a) Intra-city (b) Inter-city (c) Intra-city (d) Inter-city

**Fig. 3** Illustration of the spatial homogeneity matrix $M_{\mathcal{S}}$. The red, yellow, and blue shading in cells represent high, medium, and low spatial homogeneity values, respectively. (**a**)(**b**) Spatial homogeneity of New York. (**c**)(**d**) Spatial homogeneity of Chicago

In the link judgment, we predict that the link exists if the sigmoid connecting strength between nodes exceeds the threshold $\delta$. Our learning objective is to minimize the cross-entropy loss:

$$\min_{\mathbf{W_r},\mathbf{W_d}} -\frac{1}{(1+\omega)\,|E|}\sum_{(i,j)\in\mathcal{V}}[y\log(\sigma(\mathbf{S}(i,j))+(1-y)\log(1-\sigma(\mathbf{S}(i,j))))], \tag{6}$$

where $y=1$, $y=0$ denote the positive and negative links, $\omega=5:1$ is the ratio of negative sampling, $|E|$ is the total number of samples.

### 4.2.2 Calculation process

The calculation process is divided into two categories: intracity and intercity. In the same city, we choose three-fourths of urban regions for training to acquire the best link prediction model and then test it in the other quarter of urban regions to get the prediction scores (F1 scores) by evaluation. Notably, the F1 scores we get only belong to those quarter regions. Hence, we get the city-wide F1 scores by changing the training set and test set and finally taking the average scores. In two different cities, we transfer the best link prediction model trained from one city into all regions of another city for testing to get the city-wide F1 scores. For both calculations, we define the final city-wide F1 scores as the spatial homogeneity values of all urban regions. Higher values indicate more homogeneity. The evaluation metric F1 score is formulated as follows and the illustration of the spatial homogeneity matrix $M_{\mathcal{S}}$ is shown in Fig. 3.

$$F1=\frac{2\times Precision\times Recall}{Precision+Recall},\quad \text{where}\begin{cases}Precision=\frac{right_1}{right_1+wrong_1},\\ Recall=\frac{right_1}{right_1+wrong_0}.\end{cases} \tag{7}$$

*right* and *wrong* represent the quantities of right and wrong predictions. The inferior numbers *1* and *0* represent the positive and negative links.

## 4.3 Transfer learning for model optimization

In this subsection, we first introduce the component of model $F_\theta$ and then provide the process of model optimization under our method.

#### 4.3.1 Urban flow prediction model $F_\theta$

As with most urban computing tasks, urban flow prediction needs to consider the temporal and spatial relationship between data. Recently, GNN-based methods with temporal convolutions have attracted much interest. However, GNNs cannot be directly applied to grid-based urban flow prediction because there is no explicit graph structure. Thus, following [16, 47], we choose ST-Net as our urban flow prediction model $F_\theta$. $F_\theta$ learns the data representations of the raw urban flow data by stacking ConvLSTMs, which are the combination of CNN and LSTM. Given the urban flow $X$ based on Def.3, the output of the $l$-th ConvLSTM can be formulated as follows:

$$\mathbf{Z}^l = \text{ConvLSTM}^l(X, \theta^l_{CL}), \tag{8}$$

where $\mathbf{Z}^l$ is the data representation learned through the $i$-th layer ConvLSTM$^l$ and $\theta^l_{CL}$ denotes network parameters in this layer. The final representation $\mathbf{Z}$ encodes both the spatial and temporal information of urban flow. As a result, we obtain the final prediction through a linear layer:

$$\hat{X} = \tanh(\text{Linear}(\mathbf{Z}, \theta_L)), \tag{9}$$

where $\theta_L$ is the parameters of the linear layer. We denote $\theta = \{\theta_{CL}, \theta_L\}$ as the whole parameter set of $F_\theta$.

#### 4.3.2 Model optimization

Model optimization occupies a crucial position in urban transfer learning. We focus on the optimization of the urban flow prediction model parameter $\theta$. As shown in Pro. 2, it is realized in the source training ($\theta_\mathcal{S}$) and target fine-tuning ($\theta_\mathcal{T}$). We then improve them by weighting the source regions using spatial homogeneity:

**1) Source Training.** Same as the first urban selective transfer learning work [16], we also agree that selective source training evaluating the helpfulness of source regions is essential for urban transfer learning. Intuitively, source and target regions with similar topological structures are likely to share common temporal features. Consequently, we take the spatial homogeneity value $M_r$ as the weights of source regions to ameliorate Equation (1):

$$\theta_\mathcal{S} = \arg\min_{\theta_\mathcal{S}} \hat{\mathcal{L}_\mathcal{S}}(X_\mathcal{S}) = \sum_{r\in\mathcal{S}} \sum_{t=1}^{T_\mathcal{S}} L\left(\hat{x}_r^{(t)}, x_r^{(t)}\right) + \alpha M_r L\left(\hat{x}_r^{(t)}, x_r^{(t)}\right), \tag{10}$$

where $\alpha$ is a hyperparameter that controls the influence degree of spatial homogeneity. This optimization is to make the urban flow prediction model pay more attention to the feature learning of those more homogeneous source regions in spatial distribution.

**2) Target Fine-Tuning.** In this step, we consider the spatiotemporal similarity between source regional data and target regional data in common with previous works [40, 42]. Apart from this, it is necessary to consider the topological similarity of the regional road network between the source domain and target domain. Following the region matching strategy [40], we adopt a matching function $\mathcal{M}$: $r_\mathcal{T} \rightarrow r_\mathcal{S}$ to map each region in the target city $\mathcal{T}$ to

a suitable region in the source city $\mathcal{S}$. Our objective is to find the source region having a similar pattern in both the regional data and the regional structure with the target region. For this purpose, we calculate the correlation (i.e., $Pearson$ coefficient $P$) between each target region and source region with the corresponding data and then weight the source regions with the spatial homogeneity matrix $M_{\mathcal{S}}$ so as to match the most similar source region for each target region. The formal definition of $\mathcal{M}$ is provided as follows:

$$\begin{aligned} &\rho_{r,r^*} = M_{r^*} \cdot P(X_r, X_{r^*}), && r \in \mathcal{T}, r^* \in \mathcal{S} \\ &\rho_{r,r^*} \geq \rho_{r,r'}, && \forall r' \in \mathcal{S} \\ &\mathcal{M}(r) = r^*, && r \in \mathcal{T}, r^* \in \mathcal{S} \end{aligned} \tag{11}$$

where $X_r$ and $X_{r^*}$ represent time series of equal length $T_{\mathcal{T}}$. $\rho$ is the similarity value of the region pair and $M_{r^*}$ denotes the spatial homogeneity value in the region $r^*$. After the region matching, we then start the model optimization in this step including two optimization objectives: 1) the first objective is to minimize the prediction error described in Equation (2); 2) the second objective is to minimize the representation discrepancy between matched region pairs:

$$\min_{\theta_{\mathcal{T}}} \sum_{r \in \mathcal{T}} \sum_{t=1}^{T_{\mathcal{T}}} \rho_{r,r^*} \cdot L\left(\mathbf{z}_r^{(t)}, \mathbf{z}_{r^*}^{(t)}\right), \quad \text{where } r^* = \mathcal{M}(r). \tag{12}$$

$\mathbf{z}_r^{(t)}$, $\mathbf{z}_{r^*}^{(t)}$ are the final representations based on Equation (8). $\rho_{r,r^*}$ is the correlation value calculated between region pairs, which means that more similar region pairs will be assigned larger weights in the optimization. Finally, the two objectives are combined to form the whole optimization:

$$\theta_{\mathcal{T}} = \arg\min_{\theta_{\mathcal{T}}} \hat{\mathcal{L}}_{\mathcal{S}}(X_{\mathcal{S}}; \theta_{\mathcal{S}}) = \sum_{r \in \mathcal{T}} \sum_{t=1}^{T_{\mathcal{T}}} (1-\beta) L\left(\hat{x}_r^{(t)}, x_r^{(t)}\right) + \beta \rho_{r,r^*} \cdot L\left(\mathbf{z}_r^{(t)}, \mathbf{z}_{r^*}^{(t)}\right), \tag{13}$$

where $\beta$ is a hyperparameter meaning the weight of minimizing the representation discrepancy and minimizing the prediction error. The pseudocode of the optimization process is detailedly summarized in Algorithm 1.

## 5 Experiments

In this section, we conduct extensive experiments on four real-world spatiotemporal datasets of two cities to evaluate SHTL. To be specific, our experiments aim to answer the following questions:

- Whether the framework SHTL is effective compared with state-of-the-art transfer learning methods for urban flow prediction?
- How do the model components and hyperparameters affect the overall performance of SHTL?
- How to explain the effectiveness of spatial homogeneity value on urban transfer learning?

**Algorithm 1** Urban flow prediction model optimization.

**Input:**
Source urban flow data, $X_{\mathcal{S}}$;
Target urban flow data, $X_{\mathcal{T}}$;
Source spatial homogeneity matrix, $M_{\mathcal{S}}$;
**Output:**
Model parameter for the target domain, $\theta$.
Set $train\_epoch = 0$, $tune\_epoch = 0$;
**while** $train\_epoch < T_{train}$ **do**
  Train $\theta_{\mathcal{S}}$ on source data $X_{\mathcal{S}}$ via Equation (10) with matrix $M_{\mathcal{S}}$;
  $train\_epoch = train\_epoch + 1$;
**end while**
**while** $tune\_epoch < T_{tune}$ **do**
  Match region pairs via Equation (11) with matrix $M_{\mathcal{S}}$;
  Train $\theta_{\mathcal{T}}$ on target data $X_{\mathcal{T}}$ via Equation (13);
  $tune\_epoch = tune\_epoch + 1$;
**end while**
$\theta \leftarrow \theta_{\mathcal{T}}$; **return** $\theta$.

**Table 1** Detailed statistics of the datasets

| Dataset | Trips | City center | Time interval | Grid map | Time span |
|---|---|---|---|---|---|
| NYCTaxi | 67 million | (40.73,−73.94) | 1 h | (20,20) | 1/1,2015 – 30/6,2015 |
| NYCBike | 9 million | (40.73,−73.94) | 1 h | (20,20) | 1/1,2015 – 31/12,2015 |
| CHITaxi | 24.5 million | (41.87,−87.68) | 1 h | (20,20) | 1/1,2016 – 31/12,2016 |
| CHIBike | 9.5 million | (41.87,−87.68) | 1 h | (20,20) | 1/1,2015 – 31/12,2016 |

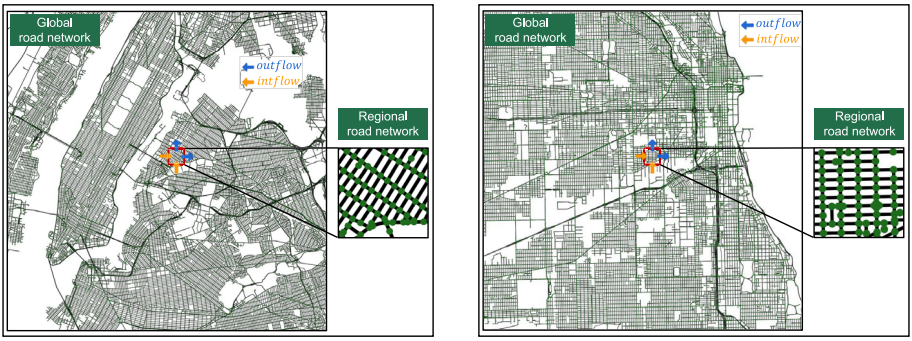
## 5.1 Experimental settings

### 5.1.1 Datasets

We collect the continuous bike and taxi trip data from New York[1] and Chicago[2] and conduct the urban flow prediction task that is widely studied in urban spatiotemporal data prediction [7, 13, 24, 26, 40]. Following previous grid-based works, each trip with a start location and end location is mapped into the inflow and outflow of the grid region. According to the latitude and longitude of the city center, we select the $0.2°(\sim 20km \times 20km)$ center areas in two cities and the area is split into $20 \times 20$ equal-size regions (i.e., each region is around $1km \times 1km$). The time interval of urban flow in each region is set as 1 hour. Given historical taxi or bike trips, urban flow prediction aims to predict the outflow and inflow of the trips in each region in the next time interval. Dataset statistics are shown in Table 1. For intracity transfer, we choose the data-abundant taxi dataset as the source domain data, and the data-scarce bike dataset as the target domain data; for intercity transfer, we choose NYCBike and CHIBike as the source domain data and the target domain data alternately. These datasets are partitioned into three sets for training, validation, and testing with a ratio of 4:1:1. More specifically, for the source domain, we split the final 20% data for testing, the 20% before data for validation, and the rest 60% data for training; for the target domain, the test and validation sets are set in

[1] https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page
https://www.citibikenyc.com/system-data.

[2] https://data.cityofchicago.org/Transportation/Taxi-Trips/wrvz-psew
https://ride.divvybikes.com/system-data.




(a) New York (b) Chicago

**Fig. 4** Visualizations of urban road networks

common with the source domain, and the training set is one month of data before validation (i.e., we suppose that the target domain has only one month of data).

In the spatial homogeneity calculation, we collect the regional road network data in New York and Chicago from OpenStreetMap.[3] Each region is matched with the region of flow data and the visualizations of the urban road networks are provided in Fig. 4. One can see that there is a "duplicate node problem" for raw urban road network data: several nodes represent the same intersection. Following [46], we merge duplicate nodes within 30 ms into a new single node and these nodes with nodes from adjacent intersections are not merged. The preprocessed regional road network features include the intersection (node) ID, the longitude and latitude of the intersection, the link between intersections, etc. For both cities, we eventually generate four kinds of spatial homogeneity matrices (a city has two matrices) as illustrated in Fig. 3. The matrices in Fig. 3a, c, b, and d are applied on NYCTaxi → NYCBike, CHITaxi → CHIBike, NYCBike → CHIBike, and CHIBike → NYCBike, respectively.

### 5.1.2 Baseline

We first compare SHTL with the following non-transfer baselines:

- **ARIMA** [36]. **Auto-Regressive Integrated Moving Average (ARIMA)** is a classic statistical time-series regression model.
- **STResNet** [50]. STResNet is a ResNet-based method for urban flow prediction, which stacks convolutional layers and residual units to capture the spatial dependency and both short- and long-term temporal dependencies.
- **ST-GFSL(DT)** [29]. We directly train ST-GFSL on limited target data.
- **ST-Net** [16]. We only train ST-Net on limited target data.

For transfer learning based baselines, we then compare with the following models:

- **Fine-Tuning** [16]. We train ST-Net on the source city with Equation (1), and then fine-tune it with Equation (2) using the data on the target city.
- **STResNet-FT**. Fine-tuning-based STResNet.
- **MetaST**. [47]. MetaST introduces meta-learning in source training (1) and transfers the spatiotemporal memory to the target city.

[3] https://www.openstreetmap.org.

- **ST-GFSL**. [29]. ST-GFSL combines gated recurrent unit (GRU) with graph attention network (GAT) to learn spatiotemporal meta-knowledge and then transfers the cross-city knowledge via parameter matching from similar meta-knowledge. Note that since ST-GFSL needs graph-based modeling, but there is not an explicit adjacency matrix in grid-based data, we construct adjacency relations by determining whether two grids are directly adjacent (including diagonal adjacency).
- **RegionTrans** [40]. After source training with Equation (1), RegionTrans computes region-wise correlation to align temporal features between source regions and target regions during target fine-tuning.
- **ST-DAAN** [42]. After source training with Equation (1), ST-DAAN confines the domain discrepancy with DAN [28] by adding adaptation layers and MMD regularization during target fine-tuning.

For fairness, both RegionTrans and ST-DAAN use ST-Net as the base urban flow prediction model. Note that due to the latitude and longitude range settings of our areas are different from CrossTReS [16] and the original point of interest (POI) and origin–destination (OD) pairs data matched with our urban regions are not provided in [16], we do not compare with it.

In addition, we propose two variants of SHTL including **SHTL-RT** (SHTL uses RegionTrans for fine-tuning) and **SHTL-STD** (SHTL uses ST-DAAN for fine-tuning) to show its compatibility.

### 5.1.3 Implementation details

The baseline methods and their parameter settings are implemented based on the original papers and official public code. Besides, we use the same settings in all baseline methods as follows for fairness. The prediction horizon $\tau = 6$ (i.e., we use the 6 previous time intervals to predict the next time intervals). The batch size is set as 24 (i.e., one day) and the learning rate is 0.001. Note that we normalize the urban flow data in each grid region into [0, 1] to facilitate the knowledge transfer between the two domains.

We implement SHTL with PyTorch. For the urban flow prediction model, the convolution kernel size is set to $3 \times 3$. Both ConvLSTM and Conv2d models contain two layers whose sizes are 64 and 2. For the link prediction model, we use the RGCN* model [46] whose layers are 3 and whose neurons are 50. The epochs and the learning rate are set as 10 and 0.001, respectively. The threshold of connection strength $\delta$ is set as 0.61.

### 5.1.4 Evaluation metrics

Two common metrics, **Root Mean Squared Error (RMSE)** and **Mean Absolute Error (MAE)**, are adopted to appraise model performances. Given predicted values $\{\hat{x}_i\}$ and ground-truth values $\{x_i\}$, the RMSE and MAE are, respectively, calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i}^{n}\left(x_i - \hat{x}_i\right)^2}, \quad \mathrm{MAE} = \frac{1}{n}\sum_{i}^{n}\left|x_i - \hat{x}_i\right|, \tag{14}$$

where n is the total number of all predicted values. Smaller metric values indicate higher accuracy. As shown in Fig. 4, there are some data-free regions in a city such as depressions and ocean regions. The same goes for urban flow data collection and hence we do not evaluate the prediction results for these regions.

**Table 2** Model comparison on bike and taxi volume prediction

| Methods | | Intracity transfer learning | | | | Intercity transfer learning | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | NYCTaxi→NYCBike | | CHITaxi→CHIBike | | NYCBike→CHIBike | | CHIBike→NYCBike | |
| | | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| No transfer | ARIMA | 0.123 | 0.0850 | 0.0678 | 0.0450 | 0.0730 | 0.0430 | 0.1020 | 0.0710 |
| | STResNet | 0.0982 | 0.0646 | 0.0663 | 0.0471 | 0.0633 | 0.0418 | 0.1064 | 0.0781 |
| | ST-GFSL(DT) | 0.0912 | 0.0592 | 0.0659 | 0.0456 | 0.0625 | 0.0429 | 0.0846 | 0.0620 |
| | ST-Net | 0.0906 | 0.0594 | 0.0657 | 0.0472 | 0.062 | 0.0424 | 0.0836 | 0.0616 |
| Transfer | Fine-Tuning | 0.0878 | 0.0569 | 0.0629 | 0.0437 | 0.0590 | 0.0400 | 0.0820 | 0.0546 |
| | STResNet-FT | 0.0934 | 0.0626 | 0.0636 | 0.0438 | 0.0612 | 0.0390 | 0.0918 | 0.0668 |
| | MetaST | 0.0880 | 0.0616 | 0.0631 | 0.0391 | 0.0602 | 0.0396 | 0.0803 | 0.0537 |
| | ST-GFSL | 0.0896 | 0.0610 | 0.0652 | 0.0443 | 0.0581 | 0.0384 | 0.0843 | 0.0559 |
| | RegionTrans | 0.0882 | 0.0583 | 0.0629 | 0.0431 | 0.0575 | 0.0404 | 0.0791 | 0.0545 |
| | ST-DAAN | 0.0870 | 0.0557 | 0.0647 | 0.0397 | 0.0589 | 0.0362 | 0.0788 | 0.0509 |
| Topological similarity-based transfer | SHTL-RT | 0.0846 | 0.0578 | 0.0613 | 0.0405 | 0.0555 | 0.0365 | 0.0774 | 0.0526 |
| | SHTL-STD | 0.0866 | **0.0551** | 0.0617 | **0.0375** | 0.0562 | **0.0329** | 0.0765 | **0.0486** |
| | SHTL | **0.0833** | 0.0563 | **0.0602** | 0.0386 | **0.0541** | 0.0358 | **0.0753** | 0.0501 |

### 5.2 Performance comparison

We measure SHTL performance on four datasets and achieve intracity and intercity transfer learning through the transfer from one dataset to another dataset. For each dataset, we report the errors of each method in Table 2. NYCTaxi → NYCBike means that NYCTaxi is selected as the source domain while NYCBike is used as the target domain. The best results are bolded and the second bests are underlined. We carry out the following observations.

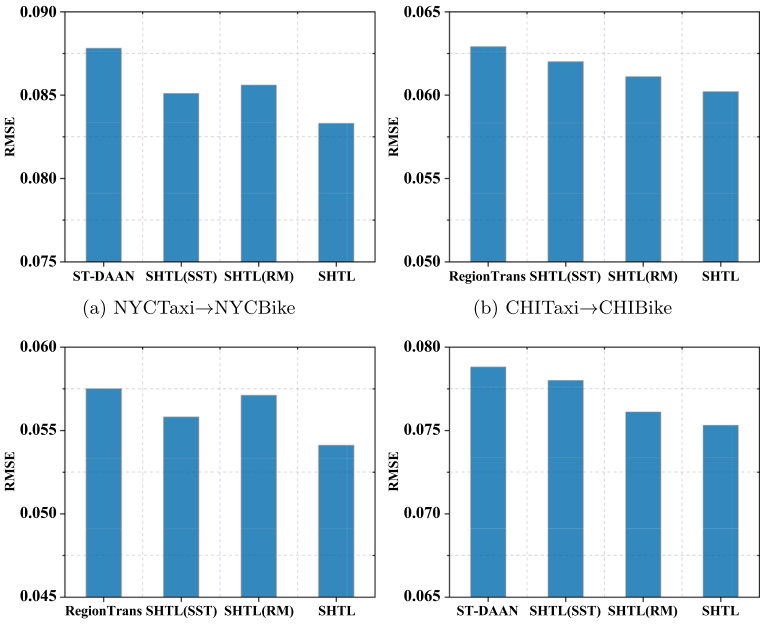
- In contrast to transfer learning methods, non-transfer models all obtain lower accuracy rates. This indicates that the problem of transferring urban flow prediction models is of pressing importance. Furthermore, ARIMA, STResNet and ST-GFSL(DT) perform worse than ST-Net, which shows the reasonableness that we use ST-Net as the base model.
- Comparing with the best baselines based on transfer learning, the framework SHTL and its variants achieve optimal performance in terms of RMSE and MAE on both urban transfer prediction ways (intracity and intercity). In particular, on NYCBike → CHIBike, our approach reduces MAE by approximately 9.1% (from 0.0362 to 0.0329) and reduces RMSE by approximately 5.9% (from 0.0575 to 0.0541). This illustrates that our method considering the spatial homogeneity of URNs is effective for urban transfer learning. In addition, the graph-based methods perform poorly, likely due to the way the adjacency matrix is constructed. Therefore, it is necessary to propose a suitable method for constructing implicit adjacency relationships to better adapt to grid-based urban flow prediction in the future.
- SHTL-RT and SHTL-STD consistently outperform their original baselines (i.e., RegionTrans and ST-DAAN), which shows that selective source training based on spatial homogeneity of SHTL can be effectively applied to general urban transfer learning methods.

### 5.3 Ablation study

We conduct ablation studies on these four datasets with SHTL(SST) (i.e., SHTL removes selective source training based on spatial homogeneity) and SHTL(RM) (i.e., SHTL removes region matching based on spatial homogeneity during target fine-tuning). Through comparing with SHTL(SST) and SHTL(RM), we test whether selective source training based on spatial homogeneity can effectively distinguish helpful source domain knowledge and test whether region matching based on spatial homogeneity can precisely match relevant source regions for target regions and eventually improve model prediction performance. As shown in Fig. 5, SHTL(SST) and SHTL(RM) consistently outperform the best baseline but their performances are lower than SHTL, which indicates that the two mechanisms we propose can better identify structurally similar regions to help urban transfer learning.

### 5.4 Parameter sensitivity analysis

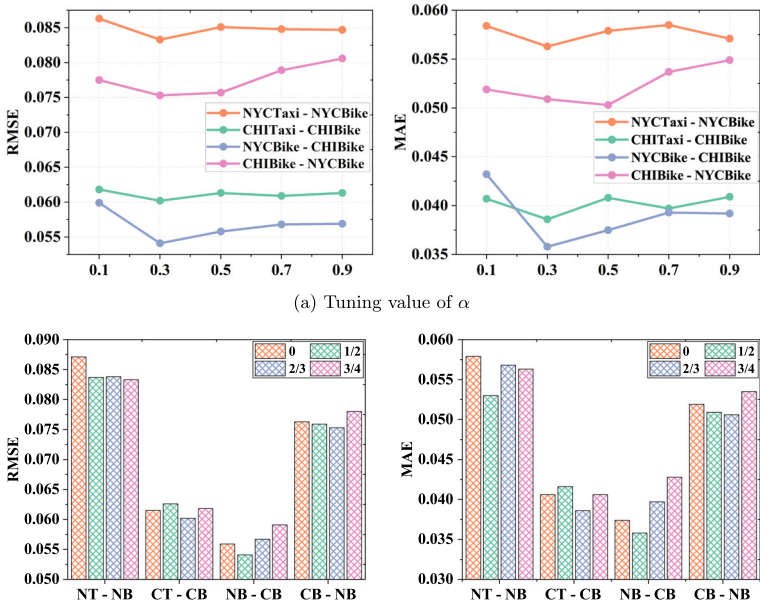
Our method contains two key hyperparameters: $\alpha$ and $\beta$. $\alpha$ controls the preference for more homogeneous regions in source training and $\beta$ balances the optimization trade-off between prediction error and representation difference. Likewise, we conduct parametric sensitivity experiments about $\alpha$ and $\beta$ on four datasets evaluated by RMSE and MAE. The results are presented in Fig. 6. Figure 6a indicates model performance variation at different settings of




(a) NYCTaxi→NYCBike (b) CHITaxi→CHIBike

(c) NYCBike→CHIBike (d) CHIBike→NYCBike

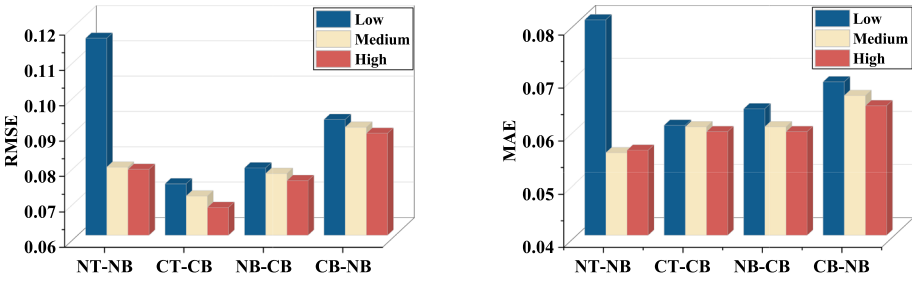
**Fig. 5** Ablation study




(a) Tuning value of $\alpha$

(b) Tuning value of $\beta$

**Fig. 6** Parameter sensitivity analysis



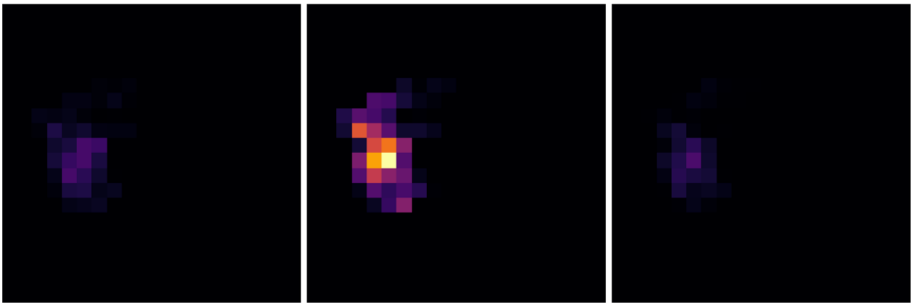

(a) RMSE with different SH value (b) MAE with different SH value

**Fig. 7** Effectiveness analysis of spatial homogeneity (SH)

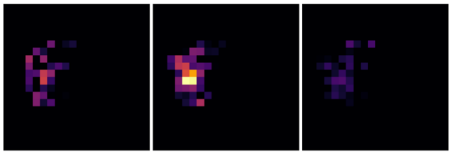
$\alpha$. One can see that almost all curves (except a curve on CHIBike $\rightarrow$ NYCBike in terms of MAE) drop rapidly and then rise slightly. This is owing to the fact that too small a $\alpha$ causes the model to ignore regions useful for transferring, while too large a $\alpha$ causes the model to pay excessive attention to the regional structure and lets more domain-specific features be transferred. Consequently, $\alpha$ is set as 0.3. The bar chart painted in Fig. 6b illustrates the effects of balance parameter $\beta$ under various settings (for brevity, NYCTaxi, NYCBike, CHITaxi, and CHIBike are simplified as NT, NB, CT, and CB, respectively). In general, the inclusion of a region matching mechanism (i.e., $\beta \neq 0$) is useful for the model. Additionally, it can be observed that when Chicago is chosen as the target domain, the model with only source selection (i.e., $\beta = 0$) achieves suboptimal performance. This means that it is also feasible to fine-tune directly without region matching in Chicago. In summary, different regional match settings should serve different types of transferring and we set $\beta$ as 0.75, 0.66, 0.5, and 0.66, respectively, to reach optimal performance.

### 5.5 Effectiveness analysis of spatial homogeneity

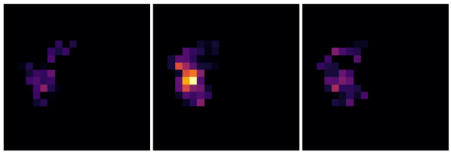
We conduct explainable region-level transfer experiments on four datasets to explain why considering spatial homogeneity is effective for urban transfer learning. Specifically, we first divide all regions of the source city into three categories based on the spatial homogeneity values: source regions with low, medium, and high spatial homogeneity values as demonstrated in Fig. 3. For each category, we randomly pick a source region to train a deep time-series prediction model and then transfer the model to all target regions for fine-tuning and testing. On account of our analysis focusing on region-level transfer, we only need to consider the influence of source regional spatial homogeneity values on each target regional time-series prediction. Here, we choose widely used LSTM with fine-tuning whose layers are 2 and hidden sizes are 64 as the deep time-series prediction model and take the average evaluation metric values testing in each target region as the result. The three-dimensional column charts displayed in Fig. 7 indicate that RMSE and MAE in the target domain decrease with increase in values of spatial homogeneity in the source regions on four transferring patterns. In other words, these source regions with higher spatial homogeneities are more helpful for region-level urban flow prediction in the target domain. Thus, our proposed SHTL which assigns higher weights to more homogenous source regions is effective.



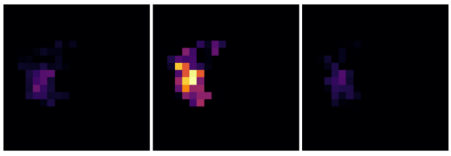
(a) Ground Truth



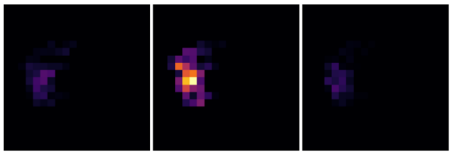
(b) ST-DAAN (NYCTaxi → NYCBike)



(c) SHTL (NYCTaxi → NYCBike)



(d) ST-DAAN (CHIBike → NYCBike)



(e) SHTL (CHIBike → NYCBike)

**Fig. 8** The case study of prediction results of ST-DAAN and SHTL in three time intervals 8:00–9:00 am (left), 16:00–17:00 pm (middle) and 22:00–23:00 pm (right)

### 5.6 Case study

To further demonstrate the effectiveness of the proposed model, we present a case study in Fig. 8. This figure displays the heat maps of predicted crowd inflows for the NYCBike dataset using the second best method ST-DAAN and our SHTL method, alongside the ground truth. Both NYCTaxi and CHIBike are chosen as source data to show intracity and intercity performances. We select three time intervals on May 30, 2015: 6:00–7:00 am, 10:00–11:00 pm, and 3:00–4:00 pm. Two of these intervals correspond to peak hours, while the other falls late at night. The urban flows during these times exhibit significant variations, with a high demand for bikes during peak hours and a notable decrease in demand during the late night. Compared to ST-DAAN, the predictions made by SHTL align more closely with the ground truth across two transfer ways, especially in the time interval 16:00–17:00 pm. This case study further indicates that our proposed SHTL method truly outperforms the ST-DAAN approach.

## 6 Conclusion

In this paper, we evaluate the framework SHTL on real-world taxi and bike data, where SHTL outperforms state-of-the-art baselines under the same settings. Two tasks including link prediction and urban flow prediction compose our proposed SHTL. The spatial homogeneity, which is defined as the F1 score evaluating link prediction performance in the first task, weights source regions to improve source training and target fine-tuning in the second task. Through these two tasks, the final urban flow prediction model considers source knowledge transferability and the similarities of both regional data and regional road networks between the source and target domain. In the future, we will apply the spatial homogeneity of URNs to other urban computing tasks (e.g., trajectory generation [19]) and fine-grained urban flow prediction [45].

**Acknowledgements** This work was supported in part by the "Pioneer" and "Leading Goose" R&D Program of Zhejiang under Grant 2023C01241, in part by the National Natural Science Foundation of China under Grant 62073295 and Grant 62072409, and in part by the Zhejiang Provincial Natural Science Foundation under Grant LR21F020003.

**Author Contributions** Yinghui Liu and Guojiang Shen made substantial contributions to the conception, acquired, analyzed, and interpreted data. Yanjie Fu revised the work critically for important intellectual content. Zehui Feng prepared figures 1-4. Zhenzhen Zhao prepared figures 5-7. All authors reviewed the manuscript.

**Data availability** The authors declare that the data are available and one can download them from: https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page https://www.citibikenyc.com/system-data https://data.cityofchicago.org/Transportation/Taxi-Trips/wrvz-psew https://ride.divvybikes.com/system-data https://www.openstreetmap.org, respectively.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. Bao X, Hu Q, Ji P et al (2022) Impact of basic network motifs on the collective response to perturbations. Nat Commun 13(1):5301
2. Barrington-Leigh C, Millard-Ball A (2020) Global trends toward urban street-network sprawl. Proc Natl Acad Sci 117(4):1941–1950
3. Boeing G (2020) A multi-scale analysis of 27,000 urban street networks: every us city, town, urbanized area, and Zillow neighborhood. Environ Plan B Urban Anal City Sci 47(4):590–608
4. Chan JYL, Bea KT, Leow SMH et al (2023) State of the art: a review of sentiment analysis based on sequential transfer learning. Artif Intell Rev 56(1):749–780
5. Chang I, Park H, Hong E et al (2022) Predicting effects of built environment on fatal pedestrian accidents at location-specific level: application of xgboost and shap. Accid Anal Prev 166:106545
6. Chiu SM, Liou YS, Chen YC et al (2023) Identifying key grid cells for crowd flow predictions based on CNN-based models with the grad-cam kit. Appl Intell 53(11):13323–13351
7. Dai G, Kong W, Liu Y et al (2023) Multi-perspective convolutional neural networks for citywide crowd flow prediction. Appl Intell 53(8):8994–9008
8. Fang Z, Wu D, Pan L, et al (2022) When transfer learning meets cross-city urban flow prediction: spatio-temporal adaptation matters. In: IJCAI'22, pp 2030–2036
9. Feng J, Lin Z, Xia T, et al (2020) A sequential convolution network for population flow prediction with explicitly correlation modelling. In: IJCAI, pp 1331–1337
10. Giacomin DJ, Levinson DM (2015) Road network circuity in metropolitan areas. Environ Plann B: Plann Des 42(6):1040–1053
11. Guo S, Lin Y, Li S et al (2019) Deep spatial-temporal 3d convolutional neural networks for traffic data forecasting. IEEE Trans Intell Trans Sys 20(10):3913–3926

12. Jin G, Liu C, Xi Z et al (2022) Adaptive dual-view wavenet for urban spatial-temporal event prediction. Info Sci 588:315–330
13. Jin G, Xi Z, Sha H et al (2022) Deep multi-view graph-based network for citywide ride-hailing demand prediction. Neurocomputing 510:79–94
14. Jin G, Liang Y, Fang Y, et al (2023a) Spatio-temporal graph neural networks for predictive learning in urban computing: a survey I. In: IEEE transactions on knowledge and data engineering
15. Jin G, Sha H, Xi Z et al (2023) Urban hotspot forecasting via automated spatio-temporal information fusion. Appl Soft Compu 136:110087
16. Jin Y, Chen K, Yang Q (2022c) Selective cross-city transfer learning for traffic prediction via source city region re-weighting. In: Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining, pp 731–741
17. Kirkley A, Barbosa H, Barthelemy M et al (2018) From the betweenness centrality in street networks to structural invariants in random planar graphs. Nat Commun 9(1):1–12
18. Kong X, Chen Q, Hou M et al (2022) Rmgen: a tri-layer vehicular trajectory data generation model exploring urban region division and mobility pattern. IEEE Trans Veh Technol 71(9):9225–9238
19. Kong X, Chen Q, Hou M, et al (2023a) Mobility trajectory generation: a survey. Artif Intell Rev pp 1–42
20. Kong X, Zhou W, Shen G et al (2023) Dynamic graph convolutional recurrent imputation network for spatiotemporal traffic missing data. Knowl Based Syst 261:110188
21. Kumar A, Singh SS, Singh K et al (2020) Link prediction techniques, applications, and performance: a survey. Physica A Stat Mech Appl 553:124289
22. Li F, Feng J, Yan H et al (2022) Crowd flow prediction for irregular regions with semantic graph attention network. ACM Trans Intell Syst Technol (TIST) 13(5):1–14
23. Li F, Yan H, Jin G, et al (2022b) Automated spatio-temporal synchronous modeling with multiple graphs for traffic prediction. In: Proceedings of the 31st ACM international conference on information & knowledge management, pp 1084–1093
24. Liang Y, Ouyang K, Sun J et al (2021) Fine-grained urban flow prediction. Proc Web Conf 2021:1833–1845
25. Liu C, Xiao Z, Wang D et al (2022) Foreseeing private car transfer between urban regions with multiple graph-based generative adversarial networks. World Wide Web 25(6):2515–2534
26. Liu C, Yang S, Xu Q, et al (2024) Spatial-temporal large language model for traffic prediction. arXiv preprint arXiv:2401.10134
27. Liu Z, Chen Y, Xia F et al (2023) Tap: traffic accident profiling via multi-task spatio-temporal graph representation learning. ACM Trans Knowl Discov Data (TKDD). https://doi.org/10.1145/3564594
28. Long M, Cao Y, Wang J, et al (2015) Learning transferable features with deep adaptation networks. In: International conference on machine learning, PMLR, pp 97–105
29. Lu B, Gan X, Zhang W, et al (2022) Spatio-temporal graph few-shot learning with cross-city knowledge transfer. In: Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining, pp 1162–1172
30. Miao H, Shen J, Cao J et al (2022) Mba-stnet: Bayes-enhanced discriminative multi-task learning for flow prediction. IEEE Trans Knowl Data Eng 35(7):7164–7177
31. Miao H, Zhao Y, Guo C, et al (2024) A unified replay-based continuous learning framework for spatio-temporal prediction on streaming data. arXiv preprint arXiv:2404.14999
32. Nigam R, Sharma DK, Jain S et al (2022) A local betweenness centrality based forwarding technique for social opportunistic IoT networks. Mob Netw Appl 27(2):547–562
33. Pinto G, Wang Z, Roy A et al (2022) Transfer learning for smart buildings: a critical review of algorithms, applications, and future perspectives. Adv Appl Energy 5:100084
34. Pournajar M, Zaiser M, Moretti P (2022) Edge betweenness centrality as a failure predictor in network models of structurally disordered materials. Sci Rep 12(1):11814
35. Schlichtkrull M, Kipf TN, Bloem P, et al (2018) Modeling relational data with graph convolutional networks. In: European semantic web conference, Springer, pp 593–607
36. Shekhar S, Williams BM (2007) Adaptive seasonal time series models for forecasting short-term traffic flow. Transp Res Rec 1:116–125
37. Strano E, Giometto A, Shai S et al (2017) The scaling structure of the global road network. R Soc Open Sci 4(10):170590
38. Tian C, Zhu X, Hu Z et al (2020) Deep spatial-temporal networks for crowd flows prediction by dilated convolutions and region-shifting attention mechanism. Appl Intell 50:3057–3070
39. Wang L, Guo B, Yang Q (2018) Smart city development with urban transfer learning. Computer 51(12):32–41
40. Wang L, Geng X, Ma X, et al (2019) Cross-city transfer learning for deep spatio-temporal prediction. In: Proceedings of the 28th international joint conference on artificial intelligence, pp 1893–1899

41. Wang S, Cao J, Philip SY (2020) Deep learning for spatio-temporal data mining: a survey. IEEE Trans Knowl Data Eng 34(8):3681–3700
42. Wang S, Miao H, Li J et al (2021) Spatio-temporal knowledge transfer for urban crowd flow prediction via deep attentive adaptation networks. IEEE Trans Intell Transp Syst 23(5):4695–4705
43. Wu Q, He K, Chen X et al (2021) Deep transfer learning across cities for mobile traffic prediction. IEEE/ACM Trans Netw 30(3):1255–1267
44. Xie Y, Xiong Y, Zhang J, et al (2023) Temporal super-resolution traffic flow forecasting via continuous-time network dynamics. Knowl Info Syst pp 1–26
45. Xu X, Wei Y, Wang P et al (2023) Diffusion probabilistic modeling for fine-grained urban traffic flow inference with relaxed structural constraint. In: ICASSP 2023–2023 IEEE international conference on acoustics. IEEE, speech and signal processing (ICASSP), pp 1–5
46. Xue J, Jiang N, Liang S et al (2022) Quantifying the spatial homogeneity of urban road networks via graph neural networks. Nat Mach Intell 4(3):246–257
47. Yao H, Liu Y, Wei Y, et al (2019) Learning from multiple cities: a meta-learning approach for spatial-temporal prediction. In: The world wide web conference, pp 2181–2191
48. Yu X, Wang J, Hong QQ et al (2022) Transfer learning for medical images analyses: a survey. Neurocomputing 489:230–254
49. Zhang J, Zheng Y, Qi D, et al (2016) Dnn-based prediction model for spatio-temporal data. In: Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems, pp 1–4
50. Zhang J, Zheng Y, Qi D (2017) Deep spatio-temporal residual networks for citywide crowd flows prediction. In: 31st AAAI conference on artificial intelligence
51. Zhou C, Gong M, Xu Z et al (2022) Urban scaling patterns for sustainable development goals related to water, energy, infrastructure, and society in China. Res Conserv Recycl 185:106443



**Yinghui Liu** is currently pursuing the PhD degree in computer science and technology with the College of Computer Science, Zhejiang University of Technology, China. His research interests include intelligent transportation systems, recommendation systems, and big data analytics.

**Guojiang Shen** received the BSc degree in control theory and control engineering and the PhD degree in control science and engineering from Zhejiang University, Hangzhou, China, in 1999 and 2004, respectively. He is currently a professor with the College of Computer Science and Technology, Zhejiang University of Technology. His current research interests include artificial intelligence theory, big data analytics, and intelligent transportation systems.

**Yanjie Fu** is an associate professor in the School of Computing and AI at the Arizona State University. He received his PhD degree from the Rutgers, the State University of New Jersey in 2016, the BE degree from the University of Science and Technology of China in 2008, and the ME degree from the Chinese Academy of Sciences in 2011. He has research experience in industry research labs, such as Microsoft Research Asia and IBM Thomas J. Watson Research Center. He has published prolifically in refereed journals and conference proceedings, such as IEEE TKDE, IEEE TMC, ACM TKDD, ACM SIGKDD, AAAI, IJCAI, VLDB, WWW, and ACM SIGIR. His research has been recognized by: 1) three junior faculty awards: US NAE Grainger Foundation Frontiers of Engineering early career engineer (2023), US NSF CAREER (2021), and NSF CRII (2018) awards; 2) several best paper (runner-up, finalist) awards (e.g., KDD18 best finalist, SIGSpatial20 best runner-up); 3) several community and industrial recognitions: 2024 Stanford Elsevier World's Top 2% Scientists, 2022 Baidu Scholar global top Chinese young scholars in AI, 2021 Aminer.org AI 2000 Most Influential Scholar Award Honorable Mention in Data Mining, 2016 Microsoft Azure Research Award; 4) several university-level awards: Reach the Stars Award, University System Research Board Award and University Interdisciplinary Research Award. He is committed to data science education. His graduated PhD students have joined academia as tenure-track faculty members. He is broadly interested in data mining, machine learning, and their interdisciplinary applications. His research aims to develop robust machine intelligence with imperfect and complex data by building tools to address framework, algorithmic, data, and computing challenges. His recent focuses are ML with geospatial, time-series data, data-centric AI (AI4data), AI for simulation and decisions, multimodal AI and foundation models), LLM. He currently serves as an associate editor of ACM Transactions on Knowledge Discovery from Data. He is a senior member of ACM and IEEE.

**Zehui Feng** received the BSc degree in Computer Science and Technology from Zhejiang University of Technology, China, in 2024. His current research interests include wireless sensor networks, machine learning, and deep learning.

**Zhenzhen Zhao** received a BSc degree in Computer Science 1and Automation from Zhejiang University of Technology, China, in 2019, and a PhD in Computer Science and Technology from Zhejiang University of Technology, China, in 2024. He is currently working at the postdoctoral research station of Zhejiang University of Technology. His research interests include graph neural networks, artificial intelligence, and big data analytics.

**Xiangjie Kong** received the BSc and PhD degrees from Zhejiang University, Hangzhou, China, in 2004 and 2009, respectively. He was an associate professor with the School of Software, Dalian University of Technology, China. He is currently a professor with the College of Computer Science and Technology, Zhejiang University of Technology, China. He has published over 200 scientific papers in international journals and conferences (with over 180 indexed by ISI SCIE). His research interests include social computing, mobile computing, and urban computing. He is a distinguished member of CCF and a member of ACM.